University of Bristol

Theory and Modelling in Chemical Sciences

Projector-Based Embedding Calculations on Deprotonation of Fluoroacetyl-CoA by Citrate Synthase

---

Author:
Xinglong Zhang

Supervisor:
Prof. Adrian Mulholland

21 March 2016 – 22 April 2016

Acknowledgements

I would like to express my sincere gratitude to Fredrick Manby and Adrian Mulholland for giving me the opportunity to work on this project at the Centre for Computational Chemistry (CCC) within the University of Bristol.

I am also grateful to Fredrick Manby, Adrian Mulholland, Marc van der Kamp, Simon Bennie and Robert Pennifold for providing excellent guidance throughout the duration of the project; their approachability and enthusiasm have been a source of inspiration. In addition, Marc provided the optimized geometries of the fluoroacetyl-CoA and oxaloacetate in enzyme citrate synthase as a starting point for this project.

I am also grateful to everyone in the CCC at Bristol for creating a hospitable environment and for making our short stay at Bristol an enjoyable one.

Abstract

Fluoroacetate is an organofluorine salt that is lethally toxic in many mammals due to its metabolism to fluorocitrate in the body catalysed by enzyme citrate synthase. Two possible stereoisomers of fluorocitrate, one of which is poisonous, could be formed depending on which proton is removed in the deprotonation step of fluoroacetyl-CoA. In this project, we combined projector-based wavefunction in density functional (WFT-in-DFT) embedding with molecular mechanics (MM) method to calculate the energy profiles of the deprotonation step in the formation of the enolate intermediates. We demonstrated that projector-based embedding is a viable technique to apply to the modelling of complex enzyme system. Coupled with its independence of the choice of DFT functionals as well as increased computational efficiency when a suitably truncated basis set is applied, this method will prove widespread in use when applying to the modelling of complex systems in the future.

1 Introduction

1.1 Enzyme Catalysis and Mechanisms

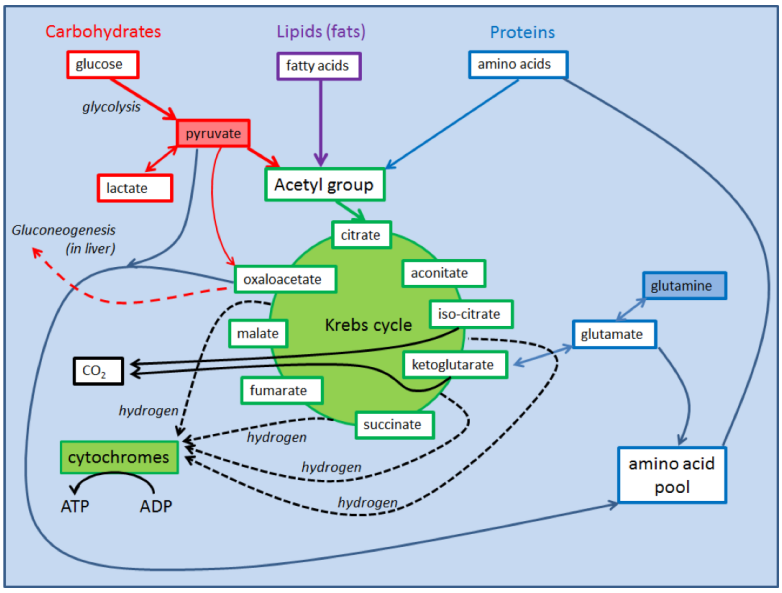
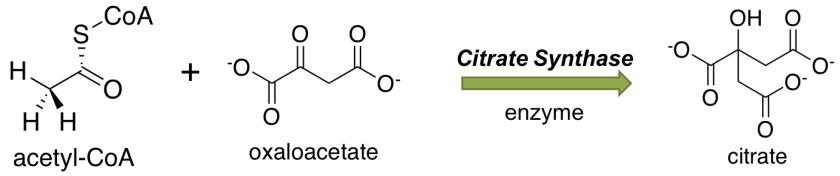


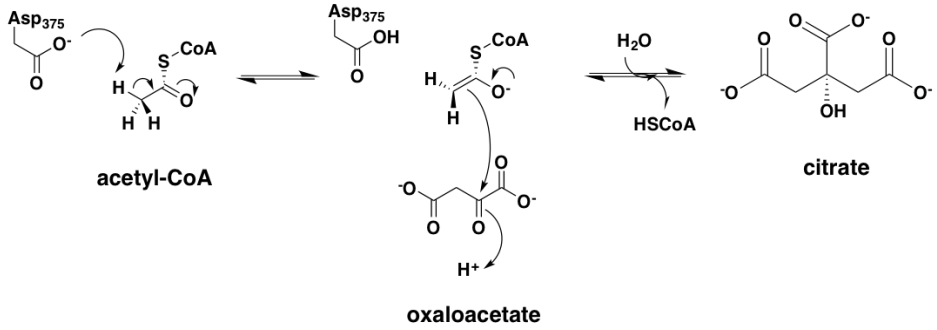
Figure 1: Schematic representation of the Citric Acid Cycle (CAC). This cycle is centrally important as it is involved in energy production, lipid metabolism as well as amino acid supply for protein synthesis. Figure adapted from electronic source.[2]

The **citric acid cycle**,[1] also known as the Krebs cycle or tricarboxylic acid cycle, consists of a series of biochemical reactions (Figure 1) involved in the production of molecular energy in all aerobic organisms. The cycle essentially oxidises acetyl-Coenzyme A (acetyl-CoA), derived variously from carbohydrates, lipids and proteins, to carbon dioxide and energy. The two-carbon (2C) acetyl-CoA first enters the citric acid cycle by reacting with four-carbon (4C) oxaloacetate to give six-carbon (6C) citrate; this conversion step is catalysed by enzyme citrate synthase:



Scheme 1: Conversion of acetyl-CoA and oxaloacetate to citrate catalysed by enzyme citrate synthase

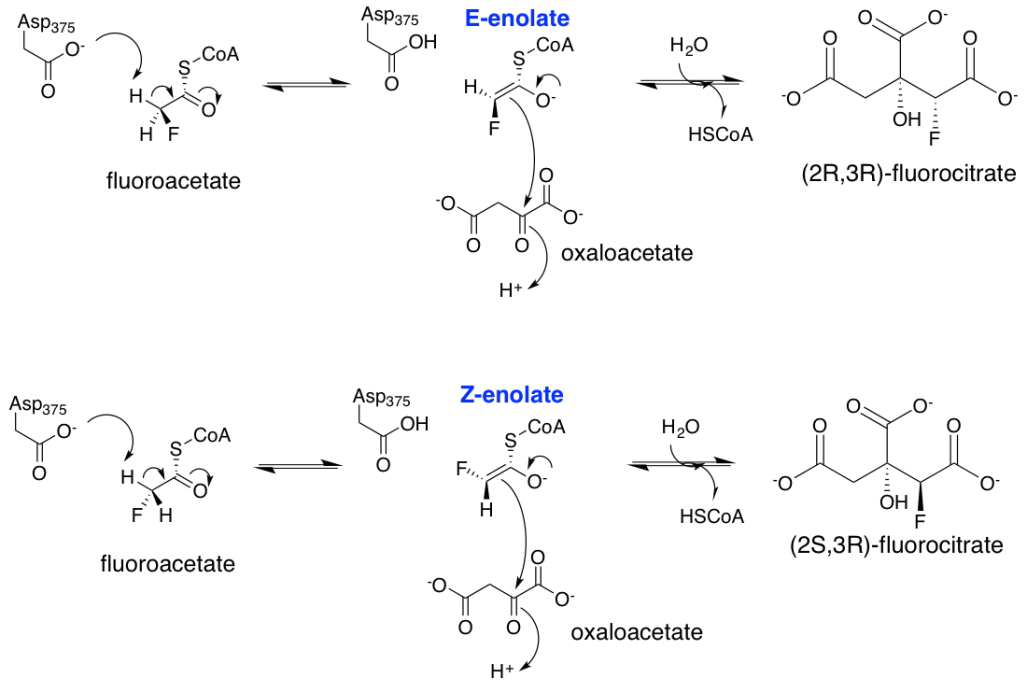
The first step of the catalysis involves a single deprotonation of methyl group on acetyl-CoA by Asp375 residue of enzyme citrate synthase, forming a stable enolate intermediate.[3,4] The negatively charged enolate oxygen is stabilised by hydrogen bond donation from His274 residue on the enzyme and a conserved water molecule.[4–6] This enolate intermediate then attacks the carbonyl carbon on oxaloacetate, giving citryl-CoA intermediate[3] which, upon hydrolysis, yields the final product of citrate. The reaction mechanism is given schematically below:



Scheme 2: Reaction mechanism for the conversion of acetyl-CoA and oxoloacetate to citrate catalysed by enzyme citrate synthase.

When fluoroacetate, an organofluorine salt that differs from acetate by the substitution of a hydrogen atom by a fluorine atom on the methyl group, is

present in the body, citrate synthase can catalyse the conversion of fluoroacetyl-CoA to fluorocitrate.[7] The enzyme mechanism is the same as in the formation of citrate from acetyl-CoA, namely a deprotonation followed by carbon-carbon condensation. However, because of the possibility to deprotonate either one of the two hydrogen atoms of the fluoroacetyl-CoA, it is now possible to form two stereoisomers of fluorocitrate (Scheme 3). The lower energy stereoisomer, (2R,3R)-fluoroacetate,[8,10] is lethally toxic; it acts as a potent inhibitor of enzyme aconitase[8,9] (enzyme catalysing the conversion of citrate to aconitate in the next step of citric acid cycle; see Figure 1), thereby halting the citric acid cycle.



Scheme 3: Conversion of fluoroacetyl-CoA and oxaloacetate to fluorocitrate catalysed by enzyme citrate synthase. Note that due to the flexibility of the C-C bond between fluoromethyl group and thioester group, two stereoisomers of fluorocitrate can be formed, depending on which hydrogen atom is deprotonated in the first step.

In this study, we are interested in the energy profiles of the deprotonation step of fluoroacetyl-CoA catalysed by citrate synthase in the formation of either E-enolate or Z-enolate.

1.2 Quantum Mechanics/Molecular Mechanics

Chemical calculations and modelling inevitably involve a compromise between speed and accuracy. Combined quantum mechanics/molecular mechanics (QM/MM) method[11–13] aims to strike a balance between these two ideals by treating a small part of a system using highly accurate but computationally demanding (*ab initio* or density functional) quantum mechanical method and the rest of the system using less accurate but computationally cheaper molecular mechanics method. The quantum mechanical methods describe the electronic structure of a subsystem, thereby allowing an accurate modelling of electron movements during bond breaking and bond formation. On the other hand, molecular mechanics method treats interactions between atoms using a potential energy function, a so-called 'force field',[14] allowing an efficient modelling of the effects of a wider environment.

Since the pivotal work by Warshel and Levitt[11] in 1976, QM/MM method have been widely applied in the field of computational enzymology,[15] where this method is used to study enzyme reactions and mechanisms. When applying QM/MM method to an enzyme system, it is important to choose and partition appropriately the regions to be treated quantum mechanically and molecular mechanically. Usually, substrate molecules and catalytic residues involved in the chemical reaction are treated using high-level QM method whereas the surrounding environment is treated by MM method. It is also important to decide on the specific QM method and MM force field to be used. There are many QM methods of varying accuracy and computational cost available. These range from fast, semiempirical[17] methods (e.g. AM1, PM3, SCC-DFTB) to the more accurate density functional[18] methods (e.g. B3LYP) and high-level, more computationally demanding (post-Hartree-Fock) *ab initio* wavefunction-based methods such as MP2,[19] CCSD(T).[20] Similarly, many choices of MM methods are available; some commonly used MM force fields include CHARMM27,[21] AMBER[22] ff99 or ff99SB and OPLS-AA.[23]

After deciding the regions for separate QM and MM treatments, it is important to decide how to treat the covalent bonds that cross the QM/MM boundary. The appropriate treatment of this interface underpins the success of QM/MM approach. Many ways exist to do so such as introducing link atoms[24–27] between QM and MM regions and using hybrid orbitals on MM atoms[28] or a 'pseudobond' to replace the covalent bond across the QM/MM boundary.[16] These methods are used to construct energy function at the interface. In addition to energy, analytical gradients of the potential energy surface are calculated in order to carry out geometry optimizations.[29]

1.3 Projector-Based WFT-in-DFT Embedding

In modelling multiscale enzyme reactions, it is computationally efficient to divide a system into subsystems where the reaction centre of an enzyme is treated using high-level quantum mechanics (QM) methods and the wider enzyme environment is treated using less expensive molecular mechanics (MM) methods (combined QM/MM[11,30]). If we look at the QM methods employed for enzyme modelling, we notice that Hartree-Fock (HF) theory can give higher reaction barriers than experiments.[5,13] Møller-Plesset second-order perturbation theory (MP2),[19] although more accurate, tends to underestimate the barriers.[13] CCSD(T) is the 'gold standard' for quantum chemical calculations.[20] However, it scales unfavourably as $O(N^7)$ where N denotes system size. Therefore, for large enzyme systems, density functional theory (DFT) offers an attractive alternative as the QM method for simulation due to its relatively lower computational cost.

One drawback with using DFT for simulations is that its accuracy is often dependent on the choice of functional employed in a particular calculation; the more accurate the functional is at describing the system, the more accurate the results are. This implies that the calculated results can vary considerably depending on the choice of functionals; it is not uncommon for DFT calculations to give qualitatively incorrect conclusions on enzyme mechanisms.[31,32] Although there is an on-going development of better functional approximations,[33–38] the lack of a systematically improvable method in DFT renders its application limited in many cases.

Wavefunction theory in density functional theory (WFT-in-DFT) embedding scheme on the QM region combines the accuracy of *ab initio* wavefunction-based technique with the speed and efficiency of density functional method. In a WFT-in-DFT embedding,[39,40] a system is divided into subsystems (denote active subsystem by A and frozen subsystem by B) where the sum of the densities of the subsystems equal to the total density of the system

$$\rho^T = \rho^A + \rho^B. \tag{1}$$

The Kohn-Sham energy in terms of the density functional is then given by

$$E[\rho^T] = E[\rho^A] + E[\rho^B] + \delta E[\rho^A, \rho^B] \tag{2}$$

where $\delta E[\rho^A, \rho^B]$ includes the non-additive Coulomb, exchange-correlation,

and kinetic energies given, respectively, by[39]

$$\Delta J[\rho^A, \rho^B] = J[\rho] - J[\rho^A] - J[\rho^B] = \int d\mathbf{r_1} \int \mathbf{r_2} \frac{\rho^A(1)\rho^B(2)}{r_{12}}; \quad (3)$$

$$\delta E_{xc}[\rho^A, \rho^B] = E_{xc}[\rho] - E_{xc}[\rho^A] - E_{xc}[\rho^B]; \quad (4)$$

$$\Delta T_s[\rho^A, \rho^B] = T_s[\rho] - T_s[\rho^A] - T_s[\rho^B]. \quad (5)$$

The non-additive kinetic energy contribution from the subsystems (Equation (5)) is the most difficult to evaluate due to the lack of an exact density functional for the kinetic energy.[39] This term has to be approximated,[41,42] resulting in errors which are largest for systems in which the subsystem densities strongly overlap.[43] Although recent development in optimized effective potential (OEP) methods[44–48] allows the exact calculation of this non-additive kinetic energy, it is computationally very demanding to do so. Therefore, projector-based embedding scheme[40] is introduced in order to negate the need to evaluate this term.

In a projector-based embedding scheme, as in all embedding schemes, the system is divided into subsystems whose densities sum up to give the total density of the whole system. Additionally, the core Hamiltonian of the active subsystem (subsystem A) embedded in frozen subsystem (subsystem B), $\mathbf{h^{A\,in\,B}}$, is modified such that the DFT-in-DFT embedding energy is the same as the full DFT energy of the total system.[39,40] A large, positive, level-shift parameter, μ, and a projector onto orbitals in subsystem B, $\mathbf{P^B}$, are introduced into the core Hamiltonian to give

$$\mathbf{h^{A\,in\,B}} = \mathbf{h} + \mathbf{J}[\rho^A + \rho^B] - \mathbf{J}[\rho^A] + \mathbf{v_{xc}}[\rho^A + \rho^B] - \mathbf{v_{xc}}[\rho^A] + \mu\mathbf{P^B} \quad (6)$$

where $\mathbf{h}$ is the original core Hamiltonian of the system; $\mathbf{J}[\rho^A + \rho^B] - \mathbf{J}[\rho^A]$ and $\mathbf{v_{xc}}[\rho^A + \rho^B] - \mathbf{v_{xc}}[\rho^A]$ give, respectively, the Coulomb and exchange correlation interactions with electrons of subsystem B. Because the level-shift parameter μ is large and positive, the net effect of applying the projector onto subsystem B is to elevate the energies of the orbitals in subsystem B so that its orbitals become inaccessible to subsystem A. In the limit $\mu \to \infty$, orbitals of subsystem A are constrained to be orthogonal to orbitals of subsystem B, so that the non-additive kinetic energy becomes incidentally zero. Perturbation theory could be applied to calculate the $\mu \to \infty$ limit to account for the unavailability of choosing μ as ∞ in a computer simulation.[39] In using orthogonal orbital sets for the subsystems, we thereby avoid the need to evaluate the non-additive kinetic energy component arising from the division of a system into subsystems.[39,40]

For enzymatic modelling using embedding scheme, it is customary to divide the whole enzyme-substrate system into three subsystems A, B and C.

Subsystems A and B (usually made up of substrates and reacting catalytic enzyme residues) are treated quantum mechanically; specifically, subsystem A is treated using high-level (correlated) wavefunction theory (WFT) whereas subsystem B is treated using more affordable DFT method. In addition, the surrounding enzyme environment (subsystem C) is treated using low-level molecular mechanics. The core Hamiltonian then has an additional term, $\mathbf{h}^{\mathbf{C}}_{\mathbf{QM/MM}}$, that describes the electronic interaction of the QM region with the MM charges in subsystem C. The core Hamiltonian in such an enzyme system is then given by

$$\mathbf{h}^{\mathbf{A\,in\,B\,in\,C}} = \mathbf{h} + \mathbf{J}[\rho^A + \rho^B] - \mathbf{J}[\rho^A] + \mathbf{v_{xc}}[\rho^A + \rho^B] - \mathbf{v_{xc}}[\rho^A] + \mu\mathbf{P}^{\mathbf{B}} + \mathbf{h}^{\mathbf{C}}_{\mathbf{QM/MM}}. \tag{7}$$

To further increase computational efficiency, a basis set truncation on subsystem A is implemented to reduce the number of two-electron integrals to calculate by reducing the number of virtual orbitals.[49] To apply this method, orbital localisation is required. This truncated embedding calculation uses *net* Mulliken populations on localised orbitals to automatically (in MOLPRO[50,51] package) screen and keep basis functions with large contributions to the density matrix of the subsystem.[49] A single truncation parameter, λ, is used to decide if a given function α should be kept or discarded;[49] it is discarded if $q_\alpha^{net} < \lambda$ where q_α^{net} is the net Mulliken population of function α. The advantage of this method is that the basis set truncation only relies on this single truncation threshold parameter λ and requires no geometry-dependent information.[49]

In summary, a typical WFT-in-DFT//MM calculation on an enzyme system involves the following steps[40,49]

1. Select a previously optimized (e.g. with QM/MM) structure of an enzyme-substrate complex.

2. Run a single Kohn-Sham DFT calculation on the QM subsystems (A + B).

3. Localise orbitals using e.g. Intrinsic Bond Orbital (IBO) localization scheme of Knizia.[52]

4. Select atoms in subsystem A for high-level QM calculations (e.g. using MP2, SCS-MP2, CCSD(T)).

5. Apply truncation parameter λ to select an atomic orbitals subset.

6. Perform high level calculations on subsystem A using the core Hamiltonian of Equation (7). This gives the WFT-in-DFT//MM energy for the whole system.

2 Computational Methods

To study the deprotonation of fluoroacetyl-CoA step in the formation the enolate intermediates, a reaction coordinate was first defined as

$$rc = d(\mathrm{C}_{\mathrm{FaCoA}}\mathrm{H}) - d(\mathrm{O}_{\mathrm{Asp375}}\mathrm{H})$$

which is the difference between the bond length of C−H of the fluoromethyl group in fluoroacetyl-CoA and the bond length of O−H between the Oxygen atom on Asp375 reside and the abstracted proton (see Figure 2, Right for reference on the chemical structures). By comparing to e.g. nudged elastic band method, this reaction coordinate was shown to accurately represent the reaction pathway.[4]

The starting point for the simulation of this reaction was prepared perviously by van der Kamp *et al.*[53] Specifically in their work, the high-resolution crystal structure of chicken citrate synthase complexed with acetyl-CoA and the inhibitor R-malate was taken[54] and acetyl-CoA was replaced by fluoroacetyl-CoA and R-malate by oxaloacetate.[31,53] Figure 2 (Left) shows a graphical representation of the model being employed. In their work, transition state conformations were generated using umbrella sampling molecular mechanics, and then iterative QM/MM optimisations at B3LYP/6-31+G(d)//CHARMM27 level were performed in both directions to obtain structures from between $rc = -1.4$Å and $rc = 1.4$Å in 0.1Å increment.[53] In the geometry optimization, Asp375 side chain, methylthioester of fluoroacetyl-CoA and oxaloacetate were chosen to be the QM region (systems A+B in Figure 2, Centre and Right). The energies for these reaction profiles were calculated at spin-component-scaled SCS-MP2 level using aug-cc-pVDZ basis set in CHARMM27 MM force field.[53] This canonical SCS-MP2//MM energies were shown to agree well with the experimental results.[53] Therefore, we use these results that they obtained as our reference for subsequent embedding calculations.

For embedding calculations, single-point WFT-in-DFT//MM calculations were performed using the Molpro 2015.1 software package.[50,51] The MM energy and QM/MM van der Waals contributions were taken from the previous B3LYP/6-31+G(d)//CHARMM27 calculations.[53] All calculations used aug-cc-pVDZ basis set for atoms in subsystems A and B. QM region (subsystems A and B) calculations were first performed using B3LYP,[55] PBE,[56] LDA[57,58] functionals as well as Hartree-Fock theory. WFT-in-DFT//MM calculations used each of these methods for subsystem B and SCS-MP2[19] for subsystem A. The atoms for subsystem A were chosen to include those directly involved in the enzyme catalytic reaction as well as those two bonds away from them.[61]

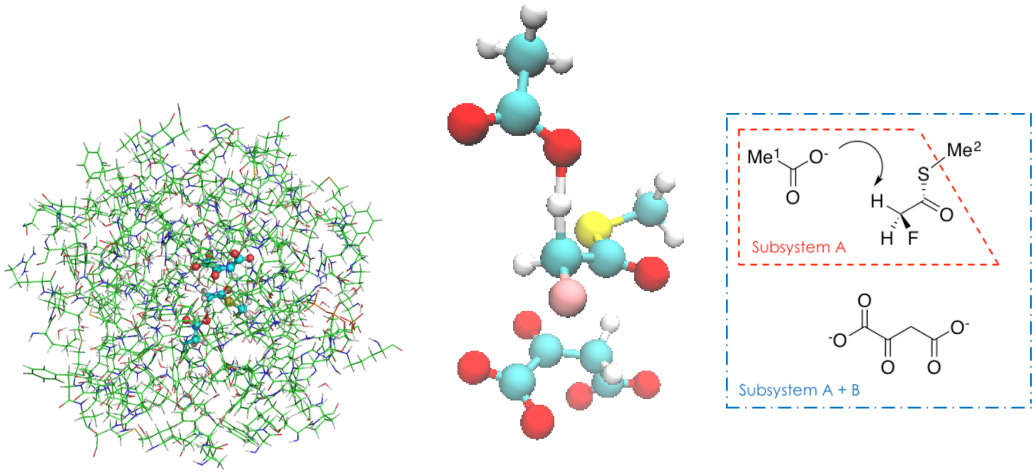


Figure 2: Left: A graphical representation of enzyme citrate synthase with fluoroacetyl-CoA and oxaloacetate substrates bound. Figure adapted from paper.[61] Centre: A ball-and-stick model of the deprotonation of fluoroacetyl-CoA by Aspartate375 residue of enzyme citrate synthase; oxaloacetate is also included. Right: Chemical structure representation of the deprotonation step of fluoroacetyl-CoA; one hydrogen atom on Me^1 side chain is treated as link atom to link the Asp375 residue to the MM region of the rest of the enzyme. Me^2 was used to replace the coenzyme A side arm in the simulation of QM region.

This region of subsystem A is shown schematically in Figure 2, Right. Subsystem A contained 80 electrons; the sulfur valency was satisfied by including the covalent bond with the methyl group. The 1s, 2s and 2p electrons from the 5 lowest-energy orbitals were treated as core electrons and were not correlated in the WFT method.[61] All other electrons, including 1s electrons on carbon, hydrogen, fluorine atoms, were correlated to achieve high accuracy. Orbital localization was achieved using the localization scheme of Knizia.[52] To apply basis-set truncation, two truncation threshold values (10^{-3} and 10^{-4}) were used. We note that the lower the truncation value, the more the functions from subsystem B are kept; if we set the threshold value to zero, then the entire basis set is retained. Visualizations of structures were done using VMD 1.9.2.[60]

3 Results and Discussion

In the formation of fluorocitrate, it was shown experimentally[62] that the minor stereoisomer, (2S,3R)-fluorocitrate, amounts to 2−3% of the major product. This corresponded to a difference in the activation energies of 2.06−2.30

kcal mol^{-1} using transition state theory. In addition, previous computational modelling[53] of the deprotonation of fluoroacetyl-CoA using QM/MM showed that SCS-MP2//MM energies corroborated well with the experimentally determined values.

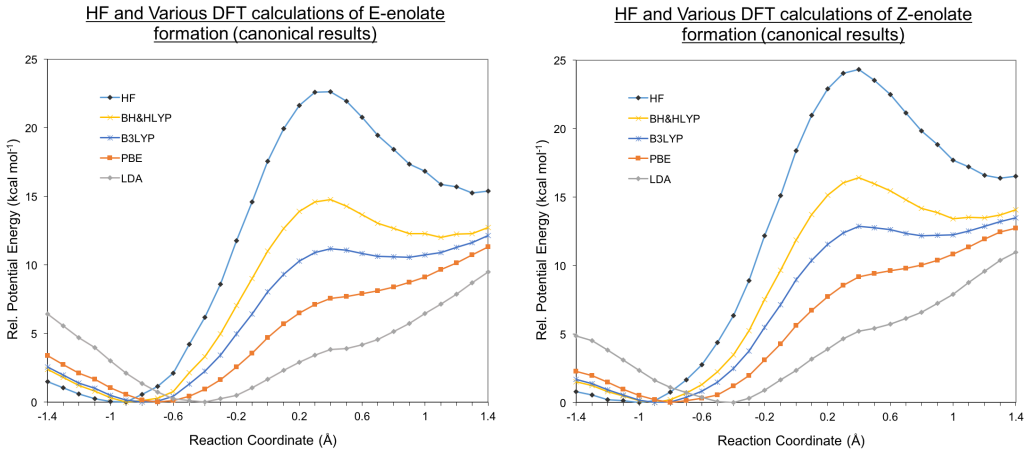


Figure 3: Energy profiles for enolate formations from canonical calculations using HF and DFT methods with full basis set. Left: E-enolate formation. Right: Z-enolate formation.

To assess the effect of choosing different density functionals on the reaction profiles of E- and Z-enolate formations, we performed a canonical calculation of the deprotonation step using Hartree-Fock and various DFT functionals on the full QM region (subsystems A + B) with full basis set. These functionals included a primitive local density approximation (LDA[57,58]), a general gradient approximation (PBE[56]), two hybrid functionals (B3LYP[55] and BH&HLYP[59]). The reaction profiles for both E- and Z-enolate formations are presented in Figure 3. All calculated energies included energy contribution from the MM region. From the figure, we can see that for both enolates, only HF and the two hybrid functionals, BH&HLYP and B3LYP, were able to predict (at least qualitatively) stable enolate formations; LDA and PBE functionals failed to predict any stable enolate formations. More importantly, we see that these different methods gave a large spread in the energy of the transition states (*ca.* 20 kcal mol^{-1}) in both enolate formations. This unequivocally demonstrated that the accuracy of a DFT calculation is strongly dependent on the choice of the functional used. We showed direct comparisons of the reaction profiles of the E- and Z-enolate formations calculated from these individual methods (except BH&HLYP) in Figure 4. Note that all methods correctly gave energies of E-enolate lower than that of Z-enolate although as expected, some functionals performed better than others

as noted above.

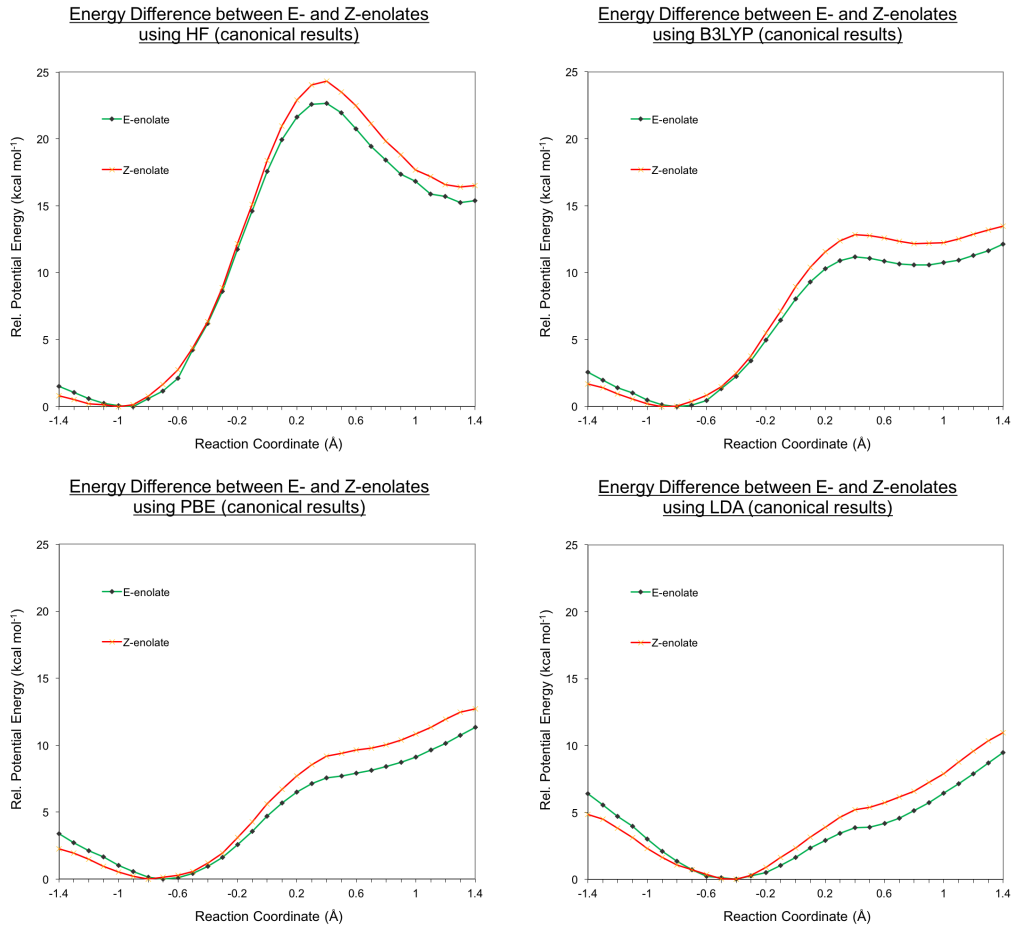


Figure 4: Energy profiles for E-enolate and Z-enolate formation from canonical calculations using HF and DFT methods on the whole QM region with full basis set. Top left: HF method. Top right: DFT B3LYP method. Bottom left: DFT PBE method. Bottom right: DFT LDA method. Note that HF and B3LYP methods predicted qualitatively correct energy profiles while PBE and LDA failed to qualitatively predict enolate formations.

Because SCS-MP2//MM energy calculated from B3LYP-optimized enzyme-substrate geometries has been shown by van der Kamp *et al*[53] to be the closest to the experimentally determined energy difference between E-enolate and Z-enolate, we will use these values as reference for all our calculations. In their study, the transition state for E-enolate formation was at an energy of 15.4 kcal mol^{-1} above the relative energy of fluoroacetyl-CoA, which was taken to be zero; the transition state for Z-enolate formation was at an energy of 17.2 kcal mol^{-1}, giving a difference of 1.8 kcal mol^{-1} in the activation energies for the formations of E-enolate and of Z-enolate. In addition, the

E-enolate intermediate was at a relative energy of 12.3 kcal mol^{-1} whereas the Z-enolate intermediate was at a relative energy of 14.3 kcal mol^{-1}, giving an energy difference of 2.0 kcal mol^{-1}; this means that the E-enolate was thermodynamically more stable than the Z-enolate.

For our embedding calculations, we note that HF-in-DFT(B3LYP, PBE, LDA) consistently gave energies higher than the reference values for both E- and Z-enolate formations (results not shown). ~~SCS-~~MP2-in-DFT(B3LYP, PBE, LDA) gave lower than the reference values for both E- and Z-enolate formations except for Z-enolate formation using ~~SCS-~~MP2-in-B3LYP embedding method (results not shown). Here, we focus on the energy profiles for E- and Z-enolate formations calculated using SCS-MP2-in-DFT(B3LYP, PBE, LDA)//MM embedding scheme.

We first showed the energy profiles for E- and Z-enolate formations separately. For each species, the reference energy profiles from earlier work[53] (canonical SCS-MP2) have been included for comparison. In addition, we also include the energy profiles from canonical DFT calculation as well as from embedding calculations using each of the two truncation parameter values. Figure 5 shows the results when using B3LYP as the density functional. We can see that in both enolates, the canonical calculations using B3LYP on the whole QM region (subsystems A + B) predicted significantly lower activation barriers. The energy profiles calculated using a basis-set truncation threshold of 10^{-4}, although about 1 kcal mol^{-1} lower, agreed quite well with the reference ones for both enolates. On the other hand, the energy profiles calculated using a higher basis-set truncation value of 10^{-3} showed an oscillatory behaviour for E-enolate formation with $rc \geq 0.6$Å and a discontinuity for Z-enolate formation at $rc = 0.3$Å. By plotting the number of basis functions used for each reaction coordinate for the embedding calculations (Figure 6), we can see that, at a higher truncation value of $\lambda = 10^{-3}$ when more basis functions were discarded, the oscillatory behaviour in the E-enolate formation correlated well with the oscillatory behaviour in the number of functions used for each reaction coordinate; the more number of basis functions used, the lower the calculated energies were. We also note that the sudden drop in the energy for the Z-enolate formation also corresponded to a sudden increase in the number of basis functions used. On the other hand, we see that at a lower truncation value ($\lambda = 10^{-4}$) when more basis functions were used, the energies were not affected, if at all, by the varying number of basis functions used for each reaction coordinate, implying that the changing number of basis functions have smaller effect on the energies calculated when more functions are included for calculations.

SCS-MP2-in-PBE and SCS-MP2-in-LDA embedding calculations gave similar results to the ones obtained for SCS-MP2-in-B3LYP embedding cal-

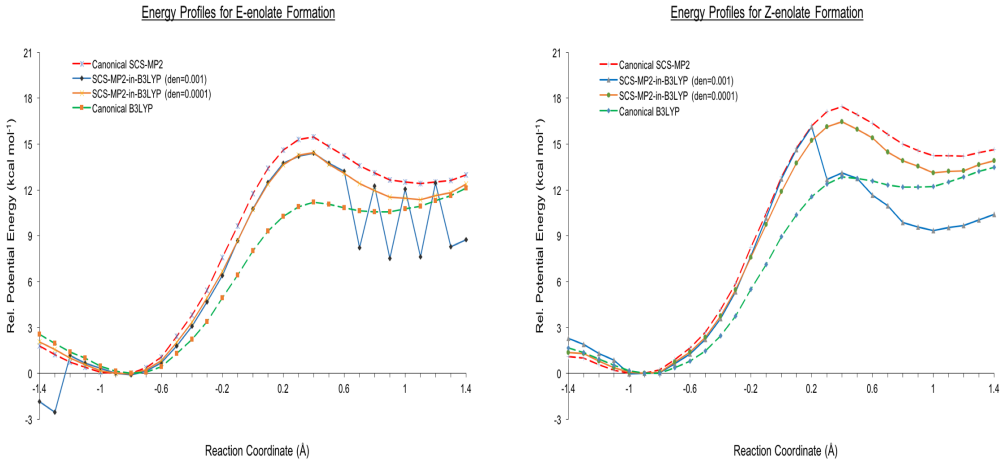


Figure 5: Energy profiles for E-enolate and Z-enolate formation from canonical SCS-MP2 (reference) and canonical B3LYP calculations using full basis set as well as SCS-MP2-in-B3LYP embedding calculations with two basis-set truncation threshold values (den=10^{-3} and den=10^{-4}). Left: E-enolate formation. Right: Z-enolate formation. Note that the energy from the MM region has been included.

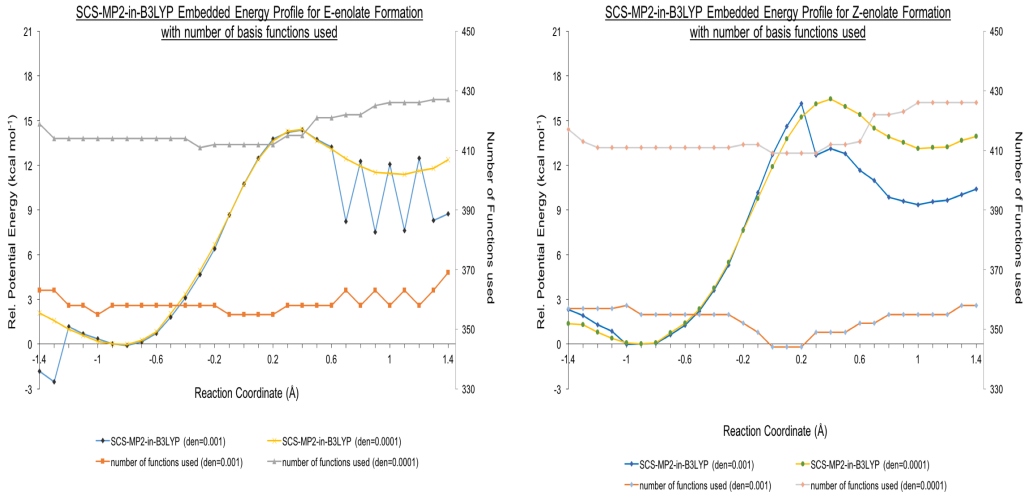


Figure 6: Energy profiles for E-enolate and Z-enolate formation from SCS-MP2-in-B3LYP embedding calculations, together with the number of basis functions used, for each of the two basis-set truncation threshold values (den=10^{-3} and den=0.0001). Left: E-enolate formation. Right: Z-enolate formation.

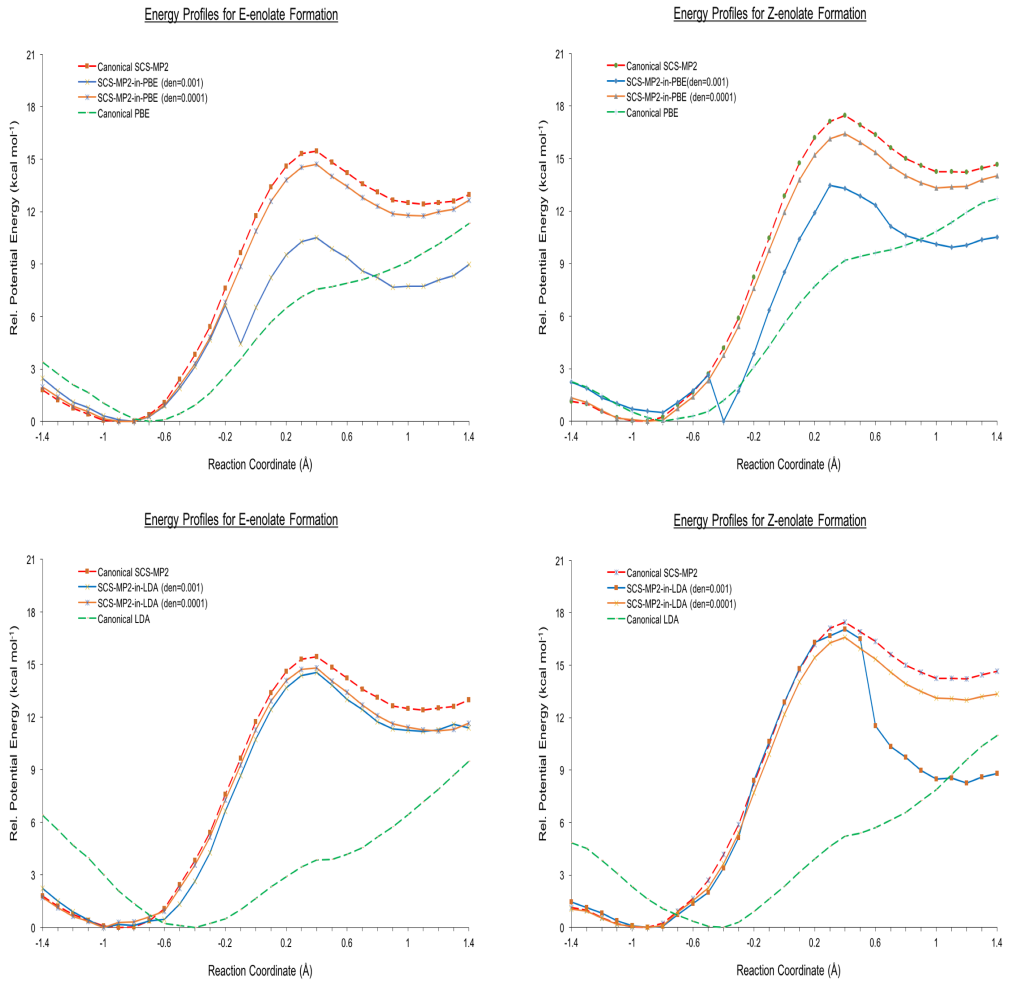


Figure 7: Energy profiles for E-enolate and Z-enolate formation from canonical SCS-MP2 (reference) and canonical PBE/LDA calculations using full basis set as well as SCS-MP2-in-PBE/LDA embedding calculations with two basis-set truncation threshold values (den=10^{-3} and den=0.0001). Note that canonical PBE and LDA results differ qualitatively from the referenced canonical SCS-MP2 results. Top Row: PBE functionals used. Bottom Row: LDA functionals used. Left: E-enolate formation. Right: Z-enolate formation. Energy from the MM region has been included.

culations. These energy profiles are shown in Figures 7. Again, we note that the energy profiles at lower basis-set truncation value ($\lambda = 10^{-4}$) agreed well with the canonical SCS-MP2 reference, although slightly lower. When more basis functions were discarded at a higher basis-set truncation value of $\lambda = 10^{-3}$, discontinuities in the reaction energy profiles occurred except for SCS-MP2-in-LDA for E-enolate formation. These discontinuities can be

shown to be similarly related to the change in the number of basis functions used for the embedding calculations (results not shown). This could be due to a substantial change in the density of the system when an important basis function was discarded. Therefore, for truncated embedding calculations, a default value of $\lambda = 10^{-4}$ is recommended to achieve an accurate result.

To return to the stereoselectivity problem of enolate formations from the deprotonation of fluoroacetyl-CoA, we compare the reaction profiles from the embedding calculations using the more accurate, smaller basis-set truncation parameter ($\lambda = 0.0001$) to the referenced canonical SCS-MP2 results. The individual SCS-MP2-in-DFT(B3LYP, PBE, LDA) profiles are included in Figure 9 in **Supplementary Data**. Here in Figure 8, we show the superimposed SCS-MP2-in-DFT embedding results for the three functionals used. Although slightly lower in energies than the referenced SCS-MP2 results, it is clear that the embedding results agreed well quantitatively amongst themselves for each enolate regardless of the choice of density functional used for subsystem B. This eliminates the dependence of the energy profile on the choice of DFT used. Notably, these WFT-embedded DFT methods yielded results significantly better than the canonical DFT (DFT on whole QM region using full basis set) results which were highly dependent on the choice of functionals (Figures 5, 7).

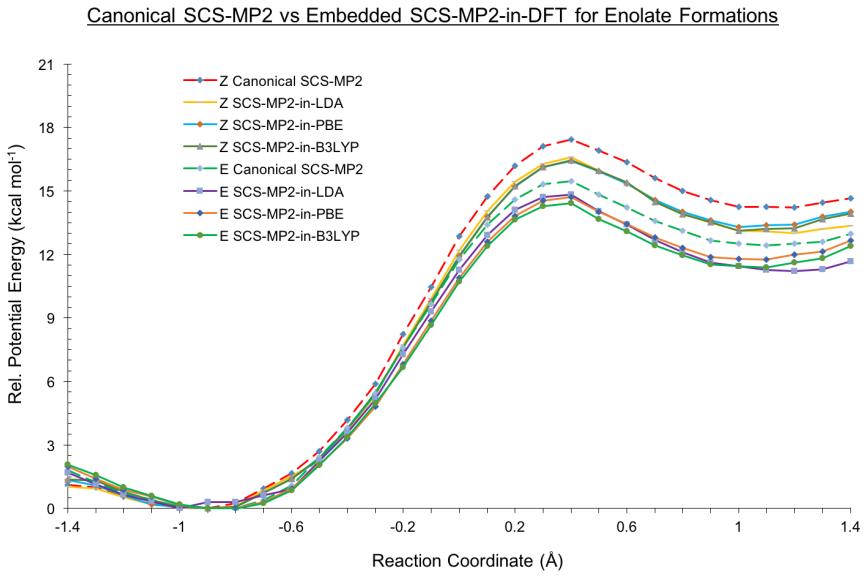


Figure 8: Energy profiles for E-enolate and Z-enolate formation from canonical SCS-MP2 (reference) and canonical LDA calculations using full basis set as well as SCS-MP2-in-LDA embedding calculations with two basis-set truncation threshold values (den=10^{-3} and den=0.0001). Left: E-enolate formation. Right: Z-enolate formation. Note that the energy from the MM region has been included.

The difference in energies between E- and Z-enolate transition states were shown using canonical SCS-MP2 to be 1.8 kcal mol^{-1}. The embedding methods gave a difference of 1.8 kcal mol^{-1} for SCS-MP2-in-B3LYP and 1.6 kcal mol^{-1} for both SCS-MP2-in-PBE and SCS-MP2-in-LDA, showing that embedding calculations can yield quantitatively correct results when we are only interested in the energy differences between transition states of different reaction pathways. The differences in energies between E- and Z-enolates from the canonical SCS-MP2 calculations were 2.0 kcal mol^{-1}. The embedding results gave differences in energies of the enolates as 2.0 kcal mol^{-1} for SCS-MP2-in-B3LYP and SCS-MP2-in-LDA and as 1.8 kcal mol^{-1} for SCS-MP2-in-PBE. Again, we note the excellent agreement of the energy differences of the enolate intermediates. Evidence from this work showed that for calculations involving relative energy differences of different intermediates/products of a (stereo-)selective reaction with multiple mechanistic pathways, embedding method could prove most computationally efficient.

4 Conclusion

We have shown that the projector-based WFT-in-DFT embedding coupled to MM is a viable method for the modelling complex systems such as enzymatic reactions. It is also possible to embed different high-level *ab initio* quantum mechanical methods in low-level density functional methods for the QM region in a wider MM environment of an enzyme surrounding. This WFT-in-DFT embedding can be easily achieved as it only requires a modification of the core Hamiltonian of an active subsystem (subsystem A) in a frozen subsystem (subsystem B). It therefore follows that different wavefunction-based embedding could be tested and systematically improved. We can therefore subsequently test the accuracy of more sophisticated methods such as CCSD(T)-in-DFT embedding.

We have also shown that a careful choice of truncation parameter for basis-set truncation can reduce the number of basis functions, thereby accelerating the computational calculations without the loss of accuracy. For default calculations, we showed that a truncation threshold of 10^{-4} is a good value to use. In addition, wavefunction embedding calculations are largely independent of the DFT functionals used for subsystem B, thereby eliminating the variability of calculated energies due to the choice of functionals in traditional canonical DFT calculations. These advantages of projector-based WFT-in-DFT embedding method, coupled with the fact that the embedding can cut across all types of bonding, will see the widespread use of this method, especially for the modelling of complex systems, in the future.

5 References

[1] Lowenstein JM. *Methods in Enzymology, Volume 13: Citric Acid Cycle.* **1969** Boston: Academic Press. ISBN 0-12-181870-5

[2] Canute's Efficient Running Site, 2007. Assessed 19 April 2016. https://canute1.wordpress.com/2013/12/05/paleo-v-high-carbohydrate-diet-the-background-to-the-evidence/

[3] Aleksandrov A., Zvereva E., Field M. *J. Phys. Chem. B* **2014**, 118, 4505−4513.

[4] van der Kamp M., Zurek J., Manby F., Harvey J., Mulholland A. *J. Phys. Chem. B*, **2010**, 114, 11303−11314.

[5] Mulholland, A. J., Lyne, P. D., Karplus, M. *J. Am. Chem. Soc.* **2000**, 122, 534−535.

[6] Yang, W., Drueckhammer, D. G. *J. Phys. Chem. B* **2003**, 107, 5986−5994.

[7] R. A. Peters, *Proc. R. Soc. London Ser. B* **1952**, 139, 143−170.

[8] Clarke D. *Neurochemical Research* **1991**, 16 (9), 1055−1058.

[9] Dumel R. J., and Kun E. *J. Biol. Chem.* **1969**, 244, 2966−2969.

[10] Stallings W. C., Monti C. T., Belvedere J. F., Preston R. K., Glusker J. P. *Arch. Biochem. Biophys.* **1980**, 203, 65−72.

[11] Warshel, A., and Levitt, M. *J. Mol. Biol.* **1967**, 103, 227−249.

[12] Senn, H. M., and Thiel, W. *Angew. Chem., Int. Ed.* **2009**, 48, 1198−1229.

[13] Lonsdale, R., Ranaghan, K. E., and Mulholland, A. J. *Chem. Commun.* **2010**, 46, 2354−2372.

[14] Mackerell, A. D. *J. Comput. Chem.* **2004**, 25, 1584−1604.

[15] Senn, H. M., and Thiel, W. *Atomistic Approaches Mod. Biol.: From Quantum Chem. Mol. Simul.* **2007**, 268, 173−290.

[16] (33) Zhang, Y. *Theor. Chem. Acc.* **2006**, 116, 43−50.

[17] Thiel, W. *WIREs Comput. Mol. Sci.* **2014**, 4, 145−157.

[18] Johnson E. R., Becke A. D. *J. Chem. Phys.* **2005**, 123, 024101.

[19] Møller, C., Plesset, M. S. *Phys. Rev.* **1934**, 46 (7), 618−622.

[20] Shavitt, I., Bartlett, R. J. *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory.* **2009**. Cambridge University Press. ISBN 978-0-521-81832-2.

[21] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. *J. Comput. Chem.* **1983**, 4, 187−217.

[22] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B. and Woods, R. J. *J. Comput. Chem.* **2005**, 26, 1668−1688.

[23] Jorgensen WL, Maxwell DS, TiradoRives J. *J. Am. Chem. Soc.* **1996**, 118, 11225−36.

[24] Maseras F, Morokuma K. *J. Comput. Chem.* **1995**, 16, 1170−79.

[25] Field MJ, Bash PA, Karplus M. *J. Comput. Chem.* **1990** 11, 700−33.

[26] Singh UC, Kollman PA. *J. Comput. Chem.* **1986**, 7:718−30.

[27] Das D, Eurenius KP, Billings EM, Sherwood P, Chatfield DC, et al. *J. Chem. Phys.* **2002**, 117, 10534−47.

[28] Gao J., Amara P., Alhambra C., Field M. J. *J. Phys. Chem. A*, **1998**, 102 (24), 4714−4721.

[29] Friesner R. A., Guallar V. *Annu. Rev. Phys. Chem.* **2005**, 56, 389−427.

[30] van der Kamp, M. W., Mulholland, A. J. *Biochemistry* **2013**, 52 (16), 2708−2728.

[31] van der Kamp, M. W., Perruccio, F., Mulholland, A. J. *Chem. Commun.* **2008**, 16, 1874−1876.

[32] Lawan, N., Ranaghan, K. E., Manby, F. R., Mulholland, A. J. *Chem. Phys. Lett.* **2014**, 608, 380−385.

[33] Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, 393 (1- 3), 51−57.

[34] Friesner, R. A.; Guallar, V. *Annu. Rev. Phys. Chem.* **2005**, 56 (1), 389−427.

[35] Grimme, S. *J. Chem. Phys.* **2006**, 124 (3), 034108.

[36] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, 120 (1-3), 215−241.

[37] Hafner, J. *J. Comput. Chem.* **2008**, 29 (13), 2044−2078.

[38]) Neese, F. *Coord. Chem. Rev.* **2009**, 253 (5-6), 526−563.

[39] Manby F. R., Stella M., Goodpaster J. D., Miller T. F. *J. Chem. Theory Comput.* **2012**, 8, 2564−2568.

[40] Goodpaster J. D.,Barnes T. A., Manby F. R., Miller T. F. *J. Chem. Phys.* **2014**, 140, 18A507.

[41] Wesolowski, T., Warshel, A. *J. Phys. Chem.* **1993**, 97, 8050−8053.

[42] Gotz, A. W.; Beyhan, S. M.; Visscher, L. *J. Chem. Theory Comput.* **2009**, 5, 3161−3174.

[43] Maseras F., Morokuma K. *J. Comp. Chem.* **1995**, 16, 1170.

[44] Goodpaster, J. D.; Ananth, N.; Manby, F. R.; Miller, T. F., III. *J. Chem. Phys.* **2010**, 133.

[45] Goodpaster, J. D.; Barnes, T. A.; Miller, T. F., III. *J. Chem. Phys.* **2011**, 134.

[46] Fux, S.; Jacob, C. R.; Neugebauer, J.; Visscher, L.; Reiher, M. *J. Chem. Phys.* **2010**, 132.

[47] Nafziger, J.; Wu, Q.; Wasserman, A. *J. Chem. Phys.* **2011**, 135.

[48] Huang, C.; Pavone, M.; Carter, E. A. *J. Chem. Phys.* **2011**, 134.

[49] Bennie S. J., Stella M., Miller T. F., Manby F. R. *J. Chem. Phys.* **2015**, 143, 024105.

[50] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schutz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Koppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pfluger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, and M. Wang, molpro, version 2012.1, a package of *ab initio* programs, **2012**, see `http://www.molpro.net`.

[51] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schutz, *WIREs: Comput. Mol. Sci.* **2012**, 2, 242.

[52] Knizia G. *J. Chem. Theory Comput.* **2013**, 9(11), 4834−4843.

[53] van der Kamp M. W., McGeagh J. D., Mulholland A. J. *Angew. Chem. Int. Ed.* **2011**, 50, 10349−10351.

[54] M. Karpusas, D. Holland, S.J. Remington. *Biochemistry.* **1991**, 30, 6024−6031.

[55] Becke, A. D. *J. Chem. Phys.* **1993**, 98 (7), 5648−5652.

[56] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, 77 (18), 3865−3868.

[57] Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, 136 (3B), B864−B871

[58] Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, 58 (8), 1200−1211.

[59] Becke A. D., *J. Chem. Phys.*, **1993**, 98,1372−77.

[60] Humphrey, W., Dalke, A. and Schulten, K. *J. Molec. Graphics*, **1996**, 14, 33−38.

[61] Bennie, S., Kamp, MVd., Pennifold, R., Stella, M., Manby, F., Mulholland, A., *J. Chem. Theory Comput.* **2016**

[62] S. Brandange, O. Dahlman, A. Mahlen, L. Mcrch, *Acta Chem. Scand. B* **1982**, 36B, 67−69.

6 Supplementary Data

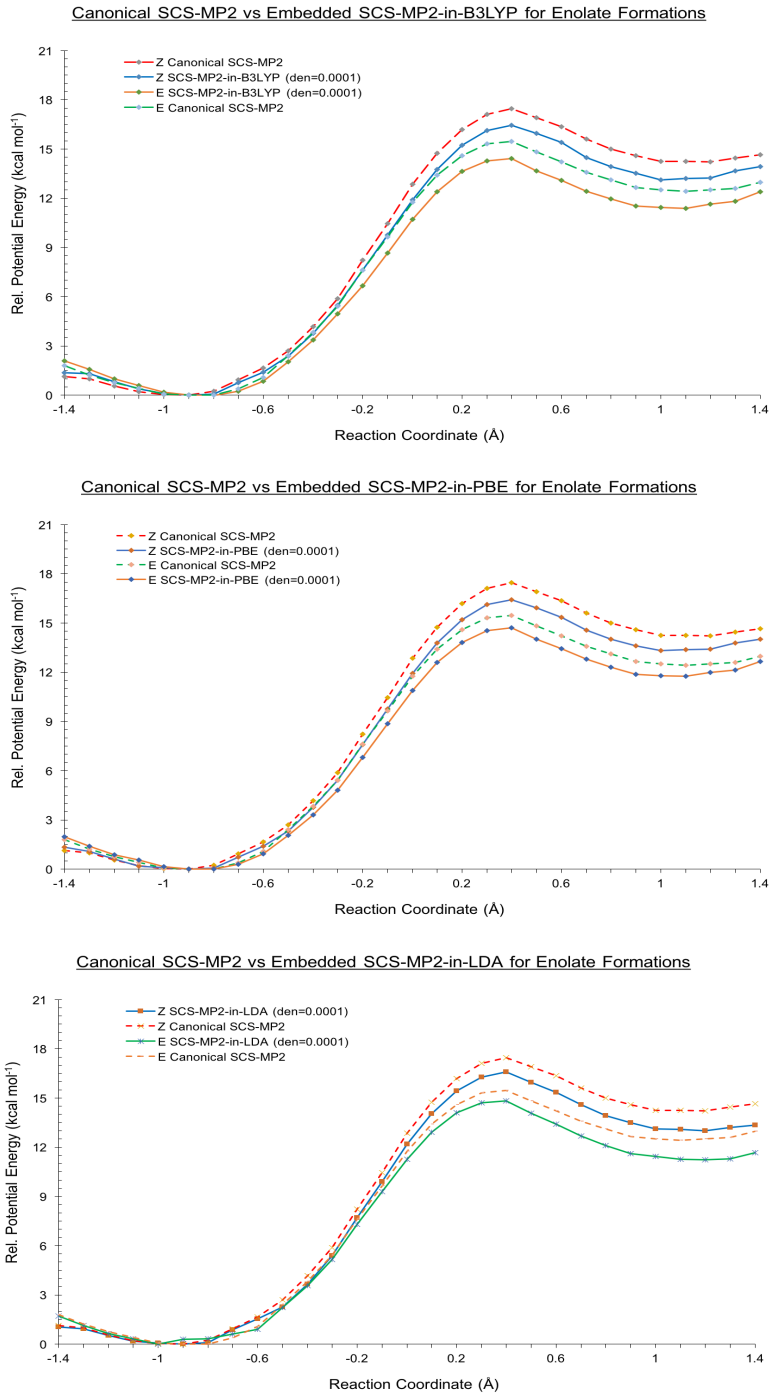


Figure 9: Energy profiles for E-enolate and Z-enolate formation from canonical SCS-MP2 (reference) and SCS-MP2-in-DFT embedding calculations using basis-set truncation threshold values of 10^{-4}. Top: SCS-MP2-in-B3LYP. Middle: SCS-MP2-in-PBE. Bottom: SCS-MP2-in-LDA.